

# Auditory time-interval perception as causal inference on sound sources

Ken-ichi Sawai<sup>1\*</sup>, Yoshiyuki Sato<sup>2</sup> and Kazuyuki Aihara<sup>1</sup>

<sup>1</sup> Institute of Industrial Science, University of Tokyo, Tokyo, Japan

<sup>2</sup> Graduate School of Information Systems, University of Electro-Communications, Tokyo, Japan

## Edited by:

Hirokazu Tanaka, Japan Advanced Institute of Science and Technology, Japan

## Reviewed by:

Hirokazu Tanaka, Japan Advanced Institute of Science and Technology, Japan  
Hiroyuki Kambara, Tokyo Institute of Technology, Japan

## \*Correspondence:

Ken-ichi Sawai, Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan.  
e-mail: [ken1@sat.t.u-tokyo.ac.jp](mailto:ken1@sat.t.u-tokyo.ac.jp)

Perception of a temporal pattern in a sub-second time scale is fundamental to conversation, music perception, and other kinds of sound communication. However, its mechanism is not fully understood. A simple example is hearing three successive sounds with short time intervals. The following misperception of the latter interval is known: underestimation of the latter interval when the former is a little shorter or much longer than the latter, and overestimation of the latter when the former is a little longer or much shorter than the latter. Although this misperception of auditory time intervals for simple stimuli might be a cue to understanding the mechanism of time-interval perception, there exists no model that comprehensively explains it. Considering a previous experiment demonstrating that illusory perception does not occur for stimulus sounds with different frequencies, it might be plausible to think that the underlying mechanism of time-interval perception involves a causal inference on sound sources: herein, different frequencies provide cues for different causes. We construct a Bayesian observer model of this time-interval perception. We introduce a probabilistic variable representing the causality of sounds in the model. As prior knowledge, the observer assumes that a single sound source produces periodic and short time intervals, which is consistent with several previous works. We conducted numerical simulations and confirmed that our model can reproduce the misperception of auditory time intervals. A similar phenomenon has also been reported in visual and tactile modalities, though the time ranges for these are wider. This suggests the existence of a common mechanism for temporal pattern perception over modalities. This is because these different properties can be interpreted as a difference in time resolutions, given that the time resolutions for vision and touch are lower than those for audition.

**Keywords:** time-interval perception, Bayesian inference, source identification, causal inference

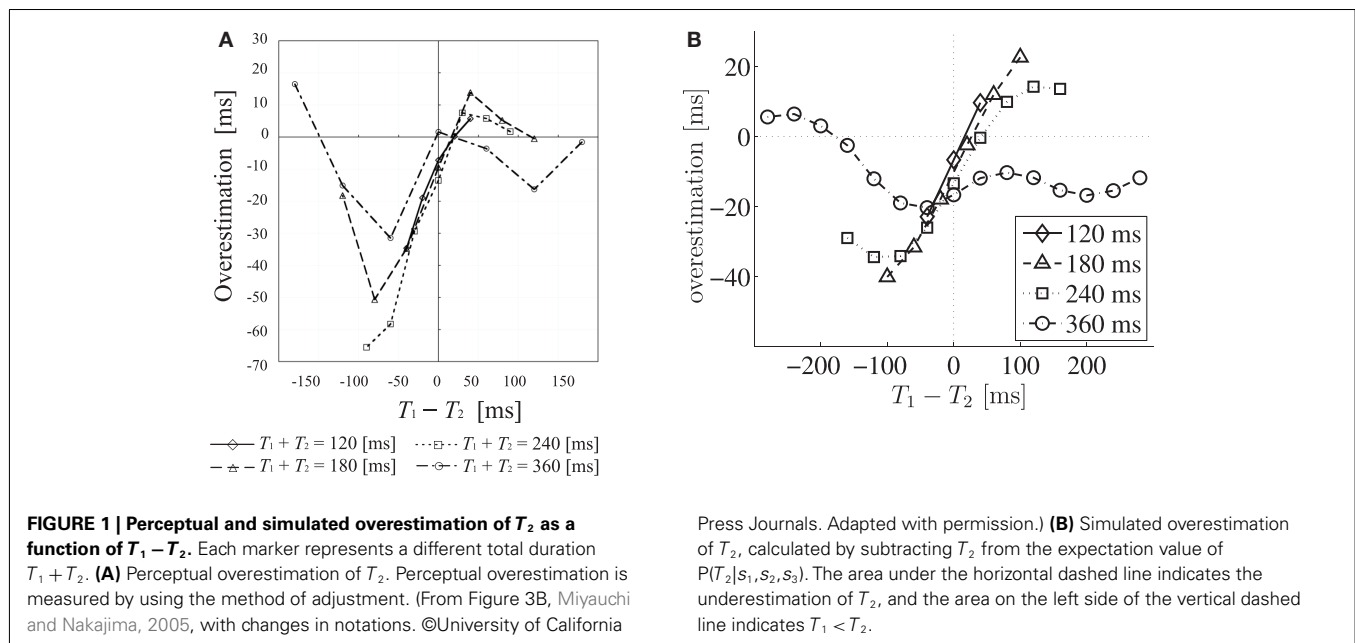
## 1. INTRODUCTION

Temporal pattern processing is necessary for all sensory modalities and these patterns contain much essential information for our brain to learn what happens in the external world. Therefore, revealing the temporal perception system is fundamental to understanding the sensory processing system, but it is not fully understood yet.

Hearing three rapid successive sounds is a good situation for investigating the time-perception system. One reason for this is that the temporal accuracy of our auditory system is higher than those for other modalities (Burr et al., 2009; Vroomen and Keetels, 2010; Occelli et al., 2011); that is, auditory experimental results reflect the actual time-perception mechanism better. In addition, a combination of two time intervals is the simplest situation of temporal pattern perception. With regard to hearing three rapid sounds on a hundred-millisecond scale, it is known that our brain sometimes misestimates the second interval depending on the relative length of the two intervals. Concretely speaking, the second interval,  $T_2$ , is perceived as shorter than the actual length in the case where  $T_2$  is equal to or a little longer than the first interval,  $T_1$ . This perceptual underestimation phenomenon was named “time-shrinking” (Nakajima et al., 1991). This illusion vanishes as the

total length  $T_1 + T_2$  increases. In addition, though the degrees of misestimation are not so large as those for the case of the time-shrinking illusion, the following phenomena on the perception of  $T_2$  have also been observed (Miyauchi and Nakajima, 2005; **Figure 1A**): overestimation of  $T_2$  when  $T_2$  is a little shorter than  $T_1$ ; underestimation of  $T_2$  when  $T_2$  is much shorter than  $T_1$ ; and overestimation of  $T_2$  when  $T_2$  is much longer than  $T_1$ . The time-shrinking illusion has been examined in other articles as well (Nakajima et al., 1992; ten Hoopen et al., 1993, 2006; Suetomi and Nakajima, 1998; Miyauchi and Nakajima, 2007; Mitsudo et al., 2009). Furthermore, it was reported that this phenomenon occurs in other sensory modalities such as visual (Arao et al., 2000) and tactile (van Erp and Spapé, 2008) senses. This fact suggests that there is a common time-perception system among sensory modalities.

A time-perception model has been proposed to explain the time-shrinking illusion (Nakajima et al., 2004). In this model, it is assumed that the subjective duration of a time-interval is proportional to the sum of the actual length and a constant length. It is also assumed that if the neural system judges the two neighboring intervals as similar, the estimating process for the latter interval is shortened and the latter interval is thus underestimated. By



these assumptions, this model can quantitatively mimic the time-shrinking illusion, namely, the underestimation of  $T_2$  caused by a shorter preceding interval  $T_1$ . However, the other misestimation phenomena when hearing three successive sounds are out of the scope of this model and cannot be reproduced by the model.

In the present study, we consider that the perceptual phenomena as mentioned above are results of effective information processing in our neural system. Sensory information, which our brain uses to infer what happens in the world, inevitably has uncertainty caused by both internal noise in our nervous system (Faisal et al., 2008) and ubiquitous fluctuation in the external world. Therefore, our brain must manage with those kinds of uncertainty, otherwise we may misunderstand the situation or regard the same experiences as different. One reasonable way for the brain to cope with the uncertainty is exploiting prior knowledge, or the experience and statistics pertaining to the situation. This strategy can be formulated by using Bayesian inference. Bayesian modeling is a powerful method for describing the human perception mechanism and has been applied to visual temporal perception (Miyazaki et al., 2005; Jazayeri and Shadlen, 2010), and more widely to human perception (Vilares and Körding, 2011, for a recent review).

## 2. MATERIALS AND METHODS

To consider the perceptual phenomena of hearing three rapid sounds, we assume a Bayesian observer who tries to solve a common source identification problem for each pair of two neighboring sounds. Further, prior to hearing, the observer assumes that sounds from the same source have short and equal intervals. The assumption of prior knowledge of short time intervals for stimuli from the same source is based on some previous works. These studies showed that the closer the two sources, the shorter are the perceived time intervals (Akerboom et al., 1983, for audition; and Goldreich, 2007; Kuroki et al., 2010, for tactile sensation). Further, with respect to the assumption of equal intervals, we can find many examples of signals aligned at almost equal intervals:

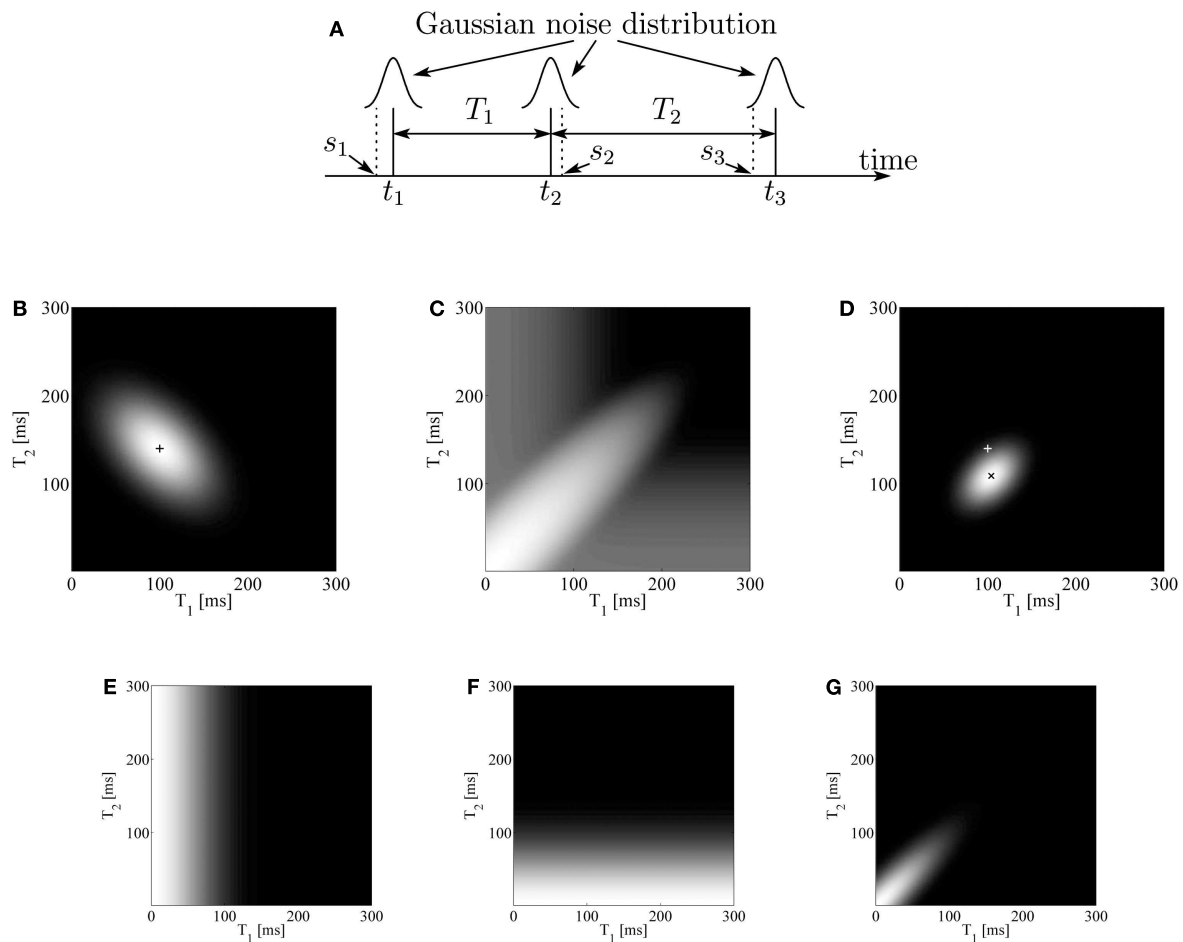
heart beats, swinging pendulum, etc. This can be because simple dynamical systems tend to generate periodical orbits, which are often observed as periodic signals generated by a limit cycle.

Here, we propose that the perception of sound intervals involves inference of causal relationship among sounds. Although there is little direct evidence for this notion, some auditory perceptual phenomenon could be associated with some form of causal judgments. For example, the time-shrinking illusion vanishes in the case wherein the temporal pattern is marked by sounds with quite different frequencies (Remijn et al., 1999). For this case, we consider that sounds with different frequencies have been judged as from independent sources. Therefore, the perceptual estimation of the latter time-interval is different from that for the case of a sound sequence composed of the same frequency. This view that sound frequency indicates source identity is also supported by an auditory psychological phenomenon (Deutsch, 1975). The perception of a common source is a kind of causal inference and should be important for making an effective inference (Körding et al., 2007; Sato et al., 2007; Shams and Beierholm, 2010). We will discuss this point further in Discussion.

Our Bayesian model assumes that our neural system cannot observe true time instants  $t_1$ ,  $t_2$ , and  $t_3$  of the sounds, but only observed times including noise  $s_1$ ,  $s_2$ , and  $s_3$ , respectively (Figure 2A). Each index of the variables indicates the order of emergence in the sound sequence. Then, our brain infers true interval durations  $T_1 = t_2 - t_1$  and  $T_2 = t_3 - t_2$  from the observation. To estimate them, our Bayesian observer composes a conditional probability, called a posterior probability,  $P(T_1, T_2 | s_1, s_2, s_3)$ . Bayesian theorem enables us to represent the posterior probability as

$$P(T_1, T_2 | s_1, s_2, s_3) = \frac{P(s_1, s_2, s_3 | T_1, T_2) P(T_1, T_2)}{P(s_1, s_2, s_3)}. \quad (1)$$

Since the denominator on the right side can be obtained by integrating the numerator over  $T_1$  and  $T_2$ , we need to consider only



**FIGURE 2 | (A)** A temporal pattern and its observation. The horizontal axis is time. Each solid vertical line indicates an actual sound timing  $t_i$ , and each dashed vertical line indicates its observed timing  $s_i$ . Each observation is made on the basis of an independent identical distribution. **(B)** Likelihood function of  $(T_1, T_2)$  given observed values of  $S_1 = 100$  ms and  $S_2 = 140$  ms (indicated as “+”). Intensity indicates the degree of likelihood. **(C)** Prior distribution of  $(T_1, T_2)$  with intensity on a logarithmic scale of the probability, illustrating that the prior takes a high value near

both axes as well as along the 45° line from the  $T_1$ -axis. **(D)** Posterior distribution of  $(T_1, T_2)$  given observed values of  $S_1 = 100$  ms and  $S_2 = 140$  ms (indicated as “+”). Intensity indicates probability. The cross sign (x) corresponds to the peak of the distribution. **(E–G)** Prior distributions of  $(T_1, T_2)$  given that **(E)** the first and second sounds come from the same source, **(F)** the second and third sounds come from the same source, and **(G)** all three sounds come from the same source. Intensity indicates probability.

terms  $P(s_1, s_2, s_3 | T_1, T_2)$  and  $P(T_1, T_2)$  in the numerator. The first term of the numerator represents how the observational values are obtained, and is formulated as

$$\begin{aligned} P(s_1, s_2, s_3 | T_1, T_2) &= \int P(s_1, s_2, s_3 | T_1, T_2, t_2) P(t_2) dt_2 \\ &\propto \int P(s_1, s_2, s_3 | t_1, t_2, t_3) dt_2 \\ &= \int P(s_1 | t_1) P(s_2 | t_2) P(s_3 | t_3) dt_2, \end{aligned} \quad (2)$$

where we assume that distribution  $P(t_2)$  is constant; knowing  $T_1$ ,  $T_2$ , and  $t_2$  is equivalent to knowing  $t_1$ ,  $t_2$ , and  $t_3$  in the second line, and the noise distributions for the timings of the three sounds are assumed to be independent from each other in the third line. We set

the distribution of the observation noise as a Gaussian distribution with the width  $\sigma_o$  and the center at a true value given as

$$P(s_i | t_i) = \frac{1}{\sqrt{2\pi}\sigma_o} \exp\left(-\frac{(s_i - t_i)^2}{2\sigma_o^2}\right), \text{ for } i = 1, 2, 3. \quad (3)$$

Here, we consider standard deviation  $\sigma_o$  to be constant with time. By substituting equation (3) into equation (2) and integrating over  $t_2$ , we obtain the following formula (see Appendix for the details of this derivation):

$$\begin{aligned} P(s_1, s_2, s_3 | T_1, T_2) &\propto \exp\left(-\frac{(T_1 - S_1)^2 + (T_2 - S_2)^2 + (T_1 - S_1)(T_2 - S_2)}{3\sigma_o^2}\right), \end{aligned} \quad (4)$$

where we introduce variables  $S_1 = s_2 - s_1$  and  $S_2 = s_3 - s_2$ , which represent the observed interval durations. Note that, given  $T_1$  and  $T_2$ ,  $t_1$  and  $t_3$  are not independent from  $t_2$  but change with  $t_2$ . Therefore, the integral range of  $t_2$  in equation (2) is  $(-\infty, \infty)$ . This term stands for the likelihood of the true intervals. Due to  $(T_1 - S_1)(T_2 - S_2)$ , this function has a negative correlation between  $T_1$  and  $T_2$ , as shown in **Figure 2B**.

Then, we formulate term  $P(T_1, T_2)$  in equation (1). This term does not relate to  $s_i (i = 1, 2, 3)$ ; that is, what our neural system has observed. Thus, this probability function represents knowledge acquired prior to the event. We model the prior knowledge of two neighboring time intervals as follows, assuming that the observer solves a source identification problem. First, our brain infers from the three successive sounds whether each pair of two neighboring sounds comes from the same source. To consider the source identification inference, we introduce variable  $C$  that represents which of the three sounds are from the same source. Here, our brain is not considered to make a judgment that the first and third sounds come from the same source while at the same time the second sound comes from another source. Thus,  $C$  represents the following four cases:

1. each sound is from an independent source,
2. the first and second sounds come from the same source and the third from another source,
3. the second and third sounds come from the same source and the first from another source,
4. all three sounds are from the same source.

Then, we assign 1, 2, 3, and 4 as the value of  $C$  to the above cases, respectively. Using the variable  $C$ , we formulate the prior distribution as

$$\begin{aligned} P(T_1, T_2) &= \sum_{C=1}^4 P(T_1, T_2, C) \\ &= \sum_{C=1}^4 P(C) P(T_1, T_2 | C). \end{aligned} \quad (5)$$

We treat the probabilities of  $C$  appearing in equation (5) as model parameters, and denote  $P(C = j) (j = 1, 2, 3, 4)$  by  $p_j$ .

Next, we formulate prior distributions  $P(T_1, T_2 | C)$  for  $C = 1, 2, 3$ , and 4, by using the assumption of equal and short intervals for sounds from the same source. The assumption is formulated as follows:

- For  $C = 1$ , there is no bias for the sound intervals. Thus, the prior distribution is a two-dimensional uniform distribution:

$$P(T_1, T_2 | C = 1) = \frac{1}{L^2}, \quad (6)$$

where  $L$  is a parameter defining the integration range.

- For  $C = 2$  and  $C = 3$ , the two sounds that come from the same source are expected to have a short interval (**Figures 2E, F**). Each prior distribution is as follows:

$$\begin{aligned} P(T_1, T_2 | C = 2) &= P(T_1 | C = 2) P(T_2 | C = 2) \\ &= \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left(-\frac{T_1^2}{2\sigma_p^2}\right) \cdot \frac{1}{L}, \end{aligned} \quad (7)$$

$$P(T_1, T_2 | C = 3) = \frac{1}{L} \cdot \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left(-\frac{T_2^2}{2\sigma_p^2}\right), \quad (8)$$

where standard deviation  $\sigma_p$  is a parameter that controls the bias toward short intervals.  $P(T_1 | C = 2)$  gives the distribution of an interval wherein the two marker sounds are from the same source, and  $P(T_2 | C = 2)$  gives the distribution of an interval wherein the two sounds come from different sources.

- For  $C = 4$ , the three markers are expected to have short and equal intervals. This distribution is expressed as a two-dimensional Gaussian distribution, with the center at the origin and a positive correlation between the two variables  $T_1$  and  $T_2$  (**Figure 2G**). Thus, this distribution can be expressed as

$$P(T_1, T_2 | C = 4) = \frac{1}{Z} \exp\left[-\left(\frac{(T_1 + T_2)^2}{2\sigma_q^2} + \frac{(T_1 - T_2)^2}{2\sigma_r^2}\right)\right], \quad (9)$$

where  $Z$  is the normalization term, and  $\sigma_q$  and  $\sigma_r$  are constant parameters. It is necessary for the prior distribution to satisfy the following condition:

$$\int P(T_1, T_2 | C = 4) dT_2 = P(T_1 | C = 2). \quad (10)$$

Given this condition, the constants  $Z$ ,  $\sigma_q$ , and  $\sigma_r$  in equation (9) are represented as follows:

$$\sigma_q^2 + \sigma_r^2 = 4\sigma_p^2, \quad (11)$$

$$Z = \pi\sigma_q\sigma_r. \quad (12)$$

New parameters  $\sigma_q$  and  $\sigma_r$  control the shape of the distribution. Since we intend the distribution to have a positive correlation between  $T_1$  and  $T_2$ ,  $\sigma_q$  should be greater than  $\sigma_r$ .

By substituting equations (6)–(9) into equation (5), we have prior distribution  $P(T_1, T_2)$ . The obtained prior distribution has a large peak at the origin of the  $T_1 - T_2$  plane, and also has high values along the  $T_1$  and  $T_2$  axes, and along the 45° line from the  $T_1$ -axis (**Figure 2C**).

Then, we obtain the posterior distribution  $P(T_1, T_2 | s_1, s_2, s_3)$  by multiplying the likelihood function of equation (4) and the prior distribution (**Figure 2D**).

### 3. RESULT

We conducted a numerical simulation to show the validity of our model. The parameter values used in the simulation are shown

in **Table 1**. There are too many parameters in our model to learn the correct values from appropriate experiments. Thus, the parameter values are chosen and adjusted so that the time scales are not strange in terms of their physical implications. For example, because the time resolution of the auditory system changes with measurement methods, a specific time resolution parameter  $\sigma_o$  cannot be decided. Therefore, we set it so that the time scale is similar to existing psychological results (Grondin and Plourde, 2007, for example). The parameter value of  $L$  is decided so as to cover the time range in which the stimuli are presented.

In this simulation, we calculated the expectation value of the marginal distribution of  $T_2$  and regarded the value as a result of the Bayesian observer's inference. Although there are some other decision-making strategies, such as maximizing the posterior probability, we chose calculating the expectation value because of its low computational cost. However, the simulation result of the maximum *a posteriori* strategy was not qualitatively different from that of the expectation value. In addition, it is yet to be ascertained which rule should be applied to a Bayesian inference (see Jazayeri and Shadlen, 2010, for this issue).

Using this simulation, our model reproduced the time-shrinking illusion; that is, the large underestimation of  $T_2$  when  $T_2$  is a little longer than  $T_1$ , due to the assumption of equal intervals. However, the amount of overestimation when  $T_2$  is a little shorter than  $T_1$  was smaller than the above underestimation. We also observed overestimation and underestimation of  $T_2$  when  $T_2$  is much longer and shorter than  $T_1$ , respectively. Moreover, our model simulation showed that the underestimation and overestimation decrease as the total length increases and that there is underestimation of  $T_2$  when  $T_2 = T_1$  (**Figure 1B**). These properties of our model were also observed in psychological experiments (**Figure 1A**).

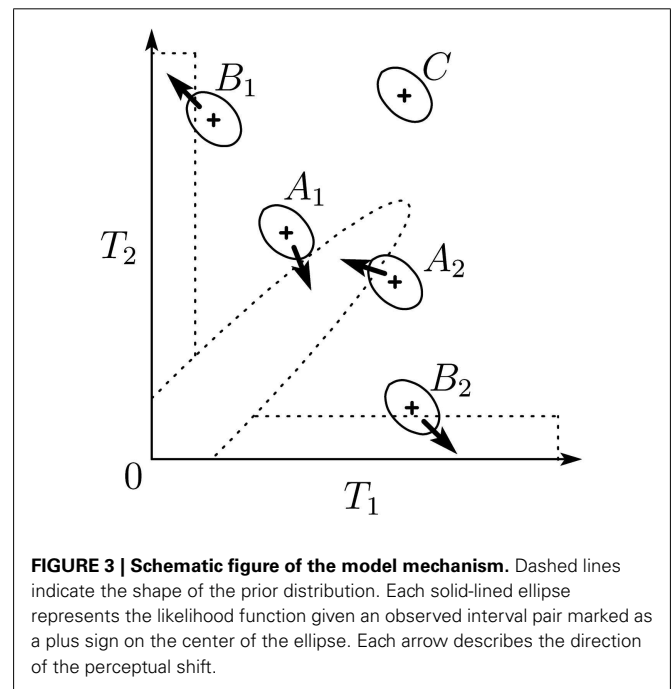
### 3.1. EXPLANATION OF THE PERCEPTION OF THREE RAPID SOUNDS

Here, we explain how our model reproduces the behavior of the human auditory system. First, when the two time intervals are similar, the observed time-interval pair stands near the diagonal line on the  $T_1$ - $T_2$  plane. Thus, the perception of three sounds shifts from noisy observation toward prior knowledge when all three sounds originate from the same source. As a result, the two intervals are perceived as more similar to each other than their observation. That is,  $T_2$  is underestimated if  $T_2$  is a little longer

than  $T_1$  (Point  $A_1$  in **Figure 3**), and  $T_2$  is overestimated if  $T_2$  is a little shorter than  $T_1$  (Point  $A_2$  in **Figure 3**). In addition, the degree of underestimation is larger than that of overestimation because the peak of the prior distribution is at the origin due to the expectation of short intervals. The expectation of short intervals also causes the underestimation of  $T_2$  when  $T_2 = T_1$ .

Next, when the intervals are dissimilar, the time-interval pair is located either near the  $T_1$ -axis or the  $T_2$ -axis on the  $T_1$ - $T_2$  plane. Therefore, perception is biased toward the  $T_1$ -axis or the  $T_2$ -axis by prior knowledge when the first two or the latter two sounds come from the same source, respectively. In addition, since the likelihood function has a negative correlation between  $T_1$  and  $T_2$ , perception shifts along the negative correlation. Thus,  $T_2$  is perceived as longer than the actual duration if  $T_2$  is longer than  $T_1$ , and vice versa (Points  $B_1$  and  $B_2$  in **Figure 3**, respectively).

In addition, the shape of the prior distribution becomes more flat as distance from the origin and the axes on the  $T_1$ - $T_2$  plane



**FIGURE 3 | Schematic figure of the model mechanism.** Dashed lines indicate the shape of the prior distribution. Each solid-lined ellipse represents the likelihood function given an observed interval pair marked as a plus sign on the center of the ellipse. Each arrow describes the direction of the perceptual shift.

**Table 1 | Parameter values in the simulation.**

Parameter	Value	Description
$\sigma_o$	25 ms	Time resolution of auditory system. The smaller this value is, the smaller is the perceptual shift
$\sigma_p$	50 ms	Strength of the bias toward short intervals. The smaller this value is, the stronger is the bias and the larger is the perceptual shift
$(\sigma_q, \sigma_r)$	(97.5, 22.2 ms)	Strength of the bias toward equal intervals. The larger the value of $\sigma_q$ is, the stronger is the bias and the larger is the perceptual shift for the case that the two intervals are similar
$(p_1, p_2, p_3, p_4)$	(0.01, 0.01, 0.01, 0.97)	Probability distribution of $C$ . The larger the value of $p_1$ is, the smaller is the perceptual shift. Increasing $p_2$ and $p_3$ results in a larger perceptual shift for the case that the two intervals are dissimilar. Increasing $p_4$ results in a larger perceptual shift for the case that the two intervals are similar
$L$	500 ms	Integration range. The shorter this value is, the smaller is the perceptual shift especially for the case that the two intervals are similar

increases. Therefore, the prior effect is weak in such areas (Point C in Figure 3).

## DISCUSSION

Our model succeeds in replicating the human perception of a simple temporal pattern. This result suggests that our brain judges the causality of sounds and expects short and equal intervals for temporal patterns in the unconscious process.

In our model, we assumed that the observer inferred the causal relationship among sounds. Although there is little evidence for this assumption, we can propose some experiments that could verify it. For example, we propose an experiment in which subjects hear three rapid sounds and report which of the three sounds come from the same source. The rate of each judgment on source identification can be predicted by calculating  $P(\text{estimated } CIT_1, T_2)$  from the present model. In addition, this experiment would also provide feedback on the parameter values of  $(p_1, p_2, p_3, p_4)$ , which are rather arbitrary in this study. By extending our model, we can also predict that the temporal pattern of sounds alters the perception of their spatial locations. Although we modeled the perception of time intervals marked by sounds in this article, we can also model the spatial perception of the sounds in almost the same form of causal inference and easily combine it with the current model. From this combined model, we predict that the same spatial patterns of sounds are perceived as spatially different if the patterns are temporally different. This is because the inference on the causal relationship among sounds is made from their temporal and spatial pattern in this model, and thus varies with temporal difference even if the actual spatial patterns are the same.

Our model has several parameters, and there exists some arbitrariness in their setting. For instance, even if we change the value of  $L$  from that in Table 1 to another value, we can reproduce a result similar to Figure 1B by adjusting parameters  $(p_1, \dots, p_4)$ . In this article, we choose quite a high value for  $p_4$  relative to the other three parameters. Although we assumed that inference was made based on observed time of sounds, in reality, we observe other features of sounds such as direction, pitch, color, volume, and so on, and all of these provide cues for the causal relationship among sounds. In the experiment we reproduced, all of these other features were kept the same for the series of three sounds, which strongly suggests that the sounds had come from the same source. We interpret  $(p_1, \dots, p_4)$  as including the cues from those other features. Thus, it might be natural to assume that  $p_4$ , which is the probability of all of the sounds coming from the same source, is considerably higher than the other possibilities. This suggests that time-interval perception depends on other sound features and, if presented with visual stimuli, also depends on visual features such as color, size, or location. In fact, it was confirmed that the result of time-interval perception differs according to the combination of stimulus pitches (Remijn et al., 1999).

## REFERENCES

- Akerboom, S., ten Hoopen, G., Olierook, P., and van der Schaaf, T. (1983). Auditory spatial alternation transforms auditory time. *J. Exp. Psychol. Hum. Percept. Perform.* 9, 882–897.
- Arao, H., Suetomi, D., and Nakajima, Y. (2000). Does time-shrinking take place in visual temporal patterns? *Perception* 29, 819–830.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Boston: The MIT Press.
- Burr, D., Silva, O., Cicchini, G. M., Banks, M. S., and Morrone, M. C. (2009). Temporal mechanisms of multimodal binding. *Proc. R. Soc. Lond. B Biol. Sci.* 276, 1761–1769.
- Deutsch, D. (1975). Two-channel listening to musical scales. *J. Acoust. Soc. Am.* 57, 1156–1160.
- Faisal, A. A., Selen, L. P. J., and Wolpert, D. M. (2008). Noise in the nervous system. *Nat. Rev. Neurosci.* 9, 292–303.

Our model could be improved by trying to replicate the experimental facts about the perception of  $T_1$ . It was reported that the direction of the perceptual shift of  $T_1$  follows the same pattern as that of  $T_2$ ; that is,  $T_1$  is underestimated when  $T_1$  is a little longer or much shorter than  $T_2$ , and  $T_1$  is overestimated when  $T_1$  is a little shorter or much longer than  $T_2$  (Miyauchi and Nakajima, 2005). This qualitative property of  $T_1$  perception can be predicted by our model. However, in that experiment, the magnitude of each perceptual shift of  $T_1$  was found to be less than that of  $T_2$ . Since the present model has symmetry between  $T_1$  and  $T_2$ , it is impossible for our model to mimic the difference between the perceptions of  $T_1$  and that of  $T_2$ . In the future, we seek to consider how we refine the present model to reproduce experimental results on the perception of  $T_1$ .

In auditory science, the issue is discerning a single sound stream in a complex of multiple sounds. This ability of the auditory system is called “auditory scene analysis” or “auditory scene segregation” (Bregman, 1990), and regarded as an important key to reveal the auditory system. Because this sound separating mechanism should involve perceptual source identification, our model may contribute to considering a sound segregation mechanism from the temporal aspect.

Finally, let us consider the time-perception mechanisms for other sensory modalities. From the psychological experiments on the visual (Arao et al., 2000) and tactile (van Erp and Spapé, 2008) time-shrinking illusions, it is known that time ranges for these modalities are broader than those for audition. The underlying reason can be understood by using the present model as follows, given that the visual and tactile time resolutions are lower than the auditory one. The perceptual bias of our model becomes weaker in a longer time scale. However, for a low-temporal-resolution modality, the perceptual bias is still relatively strong, because the observation has much uncertainty. Thus, the illusion occurs in a wider range. Though we can give a possible explanation for the difference among the modalities, the time-perception mechanisms in the sub-second scale for the other sensory modalities have not been well studied. Therefore, more research is needed before concluding that a time-perception system is shared by all sensory modalities.

## ACKNOWLEDGMENTS

The authors would like to thank Ryota Miyauchi and Yoshitaka Nakajima for their kindly providing us the high-quality image, which is used for Figure 1A. This research is supported by the Aihara Innovative Mathematical Modelling Project, the Japan Society for the Promotion of Science (JSPS) through the “Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program),” initiated by the Council for Science and Technology Policy (CSTP), and by Grant-in-Aid for Young Scientists (B) (23700309) from the Japan Society for the Promotion of Science.

- Goldreich, D. (2007). A Bayesian perceptual model replicates the cutaneous rabbit and other tactile spatiotemporal illusions. *PLoS ONE* 2, e333. doi:10.1371/journal.pone.0000333
- Grondin, S., and Plourde, M. (2007). Discrimination of time intervals presented in sequences: spatial effects with multiple auditory sources. *Hum. Mov. Sci.* 26, 702–716.
- Jazayeri, M., and Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nat. Neurosci.* 13, 1020–1028.
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE* 2, e943. doi:10.1371/journal.pone.0000943
- Kuroki, S., Watanabe, J., Kawakami, N., Tachi, S., and Nishida, S. (2010). Somatotopic dominance in tactile temporal processing. *Exp. Brain Res.* 203, 51–62.
- Mitsudo, T., Nakajima, Y., Remijn, G. B., Takeichi, H., Goto, Y., and Tobimatsu, S. (2009). Electrophysiological evidence of auditory temporal perception related to the assimilation between two neighboring time intervals. *Neuroquantology* 7, 114–127.
- Miyauchi, R., and Nakajima, Y. (2005). Bilateral assimilation of two neighboring empty time intervals. *Music Percept.* 22, 411–424.
- Miyauchi, R., and Nakajima, Y. (2007). The category of 1:1 ratio caused by assimilation of two neighboring empty time intervals. *Hum. Mov. Sci.* 26, 717–727.
- Miyazaki, M., Nozaki, D., and Nakajima, Y. (2005). Testing Bayesian models of human coincidence timing. *J. Neurophysiol.* 94, 395–399.
- Nakajima, Y., ten Hoopen, G., Hilkhuyzen, G., and Sasaki, T. (1992). Time-shrinking: a discontinuity in the perception of auditory temporal patterns. *Percept. Psychophys.* 51, 504–507.
- Nakajima, Y., ten Hoopen, G., Sasaki, T., Yamamoto, K., Kadota, M., Simons, M., et al. (2004). Time-shrinking: the process of unilateral temporal assimilation. *Perception* 33, 1061–1079.
- Nakajima, Y., ten Hoopen, G., and van der Wilk, R. (1991). A new illusion of time perception. *Music Percept.* 8, 431–448.
- Occelli, V., Spence, C., and Zampini, M. (2011). Audiotactile interactions in temporal perception. *Psychon. Bull. Rev.* 18, 429–454.
- Remijn, G., van der Meulen, G., ten Hoopen, G., Nakajima, Y., Komori, Y., and Sasaki, T. (1999). On the robustness of time-shrinking. *J. Acoust. Soc. Jpn. E* 20, 365–373.
- Sato, Y., Toyozumi, T., and Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Comput.* 19, 3335–3355.
- Shams, L., and Beierholm, U. R. (2010). Causal inference in perception. *Trends Cogn. Sci. (Regul. Ed.)* 14, 425–432.
- Suetomi, D., and Nakajima, Y. (1998). How stable is time-shrinking? *J. Music Percept. Cogn.* 4, 19–25.
- ten Hoopen, G., Hilkhuyzen, G., Vis, G., Nakajima, Y., Yamaguchi, F., and Sasaki, T. (1993). A new illusion of time perception – II. *Music Percept.* 11, 15–38.
- ten Hoopen, G., Sasaki, T., Nakajima, Y., Remijn, G., Massier, B., Rhebergen, K. S., et al. (2006). Time-shrinking and categorical temporal ratio perception: evidence for a 1:1 temporal category. *Music Percept.* 24, 1–22.
- van Erp, J. B. F., and Spapé, M. M. A. (2008). “Time-shrinking and the design of tactons,” in *Proceeding EuroHaptics '08 Proceedings of the 6th International Conference on Haptics: Perception, Devices and Scenarios*, 289–294.
- Vilares, I., and Körding, K. P. (2011). Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Ann. N. Y. Acad. Sci.* 1224, 22–39.
- Vroomen, J., and Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Atten. Percept. Psychophys.* 72, 871–884.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 July 2012; accepted: 06 November 2012; published online: 28 November 2012.

Citation: Sawai K, Sato Y and Aihara K (2012) Auditory time-interval perception as causal inference on sound sources. *Front. Psychology* 3:524. doi: 10.3389/fpsyg.2012.00524

This article was submitted to *Frontiers in Perception Science*, a specialty of *Frontiers in Psychology*.

Copyright © 2012 Sawai, Sato and Aihara. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



## APPENDIX

In this Appendix, we derive equation (4) from equations (2) and (3). First, by substituting equation (3) into equation (2), the likelihood function is written as

$$\begin{aligned}
 P(s_1, s_2, s_3 | T_1, T_2) &\propto \int_{-\infty}^{\infty} P(s_1 | t_1) P(s_2 | t_2) P(s_3 | t_3) dt_2 \\
 &= \int_{-\infty}^{\infty} \left( \frac{1}{\sqrt{2\pi}\sigma_o} \right)^3 \exp \left\{ -\frac{1}{2\sigma_o^2} [(s_1 - t_1)^2 + (s_2 - t_2)^2 + (s_3 - t_3)^2] \right\} dt_2 \\
 &\propto \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma_o^2} [(s_1 - t_2 + T_1)^2 + (s_2 - t_2)^2 + (s_3 - t_2 - T_2)^2] \right\} dt_2.
 \end{aligned} \tag{A1}$$

Then, by introducing variables  $S_1 = s_2 - s_1$  and  $S_2 = s_3 - s_2$ , and substituting  $u$  for  $t_2 - s_2$ , we obtain equation (4) as follows:

$$\begin{aligned}
 P(s_1, s_2, s_3 | T_1, T_2) &\propto \int_{-\infty}^{\infty} \exp \left( -\frac{1}{2\sigma_o^2} \{ [(s_2 - S_1) - t_2 + T_1]^2 + (s_2 - t_2)^2 + [(s_2 + S_2) - t_2 - T_2]^2 \} \right) dt_2 \\
 &= \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma_o^2} [(u - T_1 + S_1)^2 + u^2 + (u + T_2 - S_2)^2] \right\} du \\
 &= \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma_o^2} [3u^2 - 2(T_1 - T_2 - S_1 + S_2)u + (T_1 - S_1)^2 + (T_2 - S_2)^2] \right\} du \\
 &= \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma_o^2} \left[ 3 \left( u - \frac{T_1 - T_2 - S_1 + S_2}{3} \right)^2 \right] \right\} du \\
 &\quad \times \exp \left\{ -\frac{1}{2\sigma_o^2} \left[ -\frac{(T_1 - T_2 - S_1 + S_2)^2}{3} + (T_1 - S_1)^2 + (T_2 - S_2)^2 \right] \right\} \\
 &\propto \exp \left[ -\frac{(T_1 - S_1)^2 + (T_2 - S_2)^2 + (T_1 - S_1)(T_2 - S_2)}{3\sigma_o^2} \right].
 \end{aligned} \tag{A2}$$